Exploring the usability of sound strategies for guiding task: toward a generalization of sonification design

Gaëtan Parseihian, Charles Gondre, Mitsuko Aramaki, Richard Kronland-Martinet, Sølvi Ystad

Exploring the usability of sound strategies for guiding task: toward a generalization of sonification design

Gaëtan Parseihian[1], Charles Gondre[1], Mitsuko Aramaki[1], Richard Kronland Martinet[1], and Sølvi Ystad[1] ⋆

LMA, CNRS, UPR 7051, Aix-Marseille Univ, Centrale Marseille
name@lma.cnrs-mrs.fr

Abstract. This article aims at providing a new Parameter Mapping Sonification approach in order to facilitate and generalize sonification design for different applications. First a definition of the target as a concept that enables a general sonification strategy that is not limited to specific data types is given. This concept intends to facilitate the separation between sound and information to display. Rather than directly displaying data dimensions through the variation of a specific sound parameter, the approach aims at displaying the distance between a given data value and the requested value. Then a taxonomy of sound strategies based on sound that allow the construction of several strategy types is presented. Finally, several sound strategies are evaluated with a user experiment and the taxonomy is discussed on the basis of user's guidance behavior during a guiding task.

1 Introduction

Thanks to the development of research in data processing and auditory display, the use of sound in user interfaces has considerably grown over the past few decades. Employed as a means of conveying information, auditory display exploits the superior ability of the human auditory system to recognize temporal changes and patterns. It has the advantage of relieving the visual system when it is overcrowded with information or busy with another task or allows to supply it when it is not available (if the user is physically visually impaired or as a result of environmental factors such as smoke). Furthermore, our ability to monitor and process multiple auditory data stream at the same time and our ability for rapid auditory detection place the auditory modality as a good candidate to supply or augment the vision.

Using sound to communicate information has been called sonification. In [14], Kramer defines sonification as "the use of non-speech audio to convey information or perceptual data". Many studies have investigated methods of conversion from data to sound depending of the type of information. From *auditory icons*

⋆ This work is supported by the MetaSon Project - CONTINT 2010: ANR-10-CORD-010 - French National Research Agency (ANR).

and *earcons* that are brief sounds used to monitor events in user interfaces, to *parameter mapping sonification*, different sonification techniques have been defined. The Sonification Handbook [11] provides a good introduction to various methods depending on the application.

This article is based on the Parameter Mapping Sonification (PMS) approach. This method consists in representing changes in data dimension through an auditory variation [13,20,10]. Since its first definition, a number of softwares have been constructed to help the engineer to design PMS for different applications [2,4,5,6,7,12,15,17]. These softwares mostly use the basic sound parameters (frequency, intensity and tempo) as the principal mapping parameters applied on sound synthesis or MIDI instruments. Other parameters like right-left panning [12,5,7], timbre (proposing the choice of a MIDI instrument [12,15,7], or allowing to control the frequency shape or the brightness of the sound [2,16]), its rhythm [15] or the time gap between sounds, the consonance and the register [12,17] are also frequently encountered.

For instance, in [21], Walker and Kramer explored the influence of the sound parameter choice and the effect of the mapping polarity on a task of generic process control in a widget factory. In this study, the authors monitored four data types from the fictitious factory (temperature, pressure, size and rate) with different auditory dimensions (loudness, pitch, tempo and onset time) while changing mapping strategies and polarity for each experimental group. By measuring response time and precision, they showed a strong influence of the mapping strategies and the mapping polarity: some data types are best represented by a particular display dimension with a specific polarity (loudness is, for example the best dimension to represent temperature and tempo is not necessarily the best choice to represent the rate).

Since its first conceptualization by [14], many works have studied PMS for specific data dimensions (pressure, size, temperature, number of dollars, urgency, danger, ...). Since they require awareness of the user's expectations, the results of these studies are quasi impossible to generalize for the design of other applications. In practice, the PMS softwares usually allow the designer to choose different mapping strategies for the sonification of several multidimensional data types, but none intend to help the designer on the choice of the sound strategies, the optimal polarity for these mappings and the scaling of the parameters. In addition, the scaling function between display dimension and data change also appears to be an important part of the design. However, it is difficult to find information about it in sonification literature. Few studies aim at exploring its influence on the mapping efficiency [18,19] and most of the works doesn't mention the range values of the studied display parameters. To face this problem, Grond and Berger [10] purpose to establish this function while taking into account the perception of each sounds parameter, their Just Noticeable Difference (JND) and the masking threshold, and Ferguson *et al.* [9] suggest to use psychoacoustic models for the sonification. They introduce the psychoacoustic definition of the pitch, the loudness, the roughness and the brightness and propose a theoretical implementation of these parameters. If the perception of some of these parameters is well known (the JND of the pitch and the loudness have been defined several years ago), it is difficult to find a consensus for the JND perception of the roughness and the brightness.

The present article aims at generalizing PMS strategies for guiding tasks by identifying perceptual cues that are optimally efficient for precise or rapid guidance. It first introduces the notion of "target" and define a normalized distance between a given data value and the requested data value. This notion allow to disconnect the sonification from the application and the type of data to display. It proposes then a taxonomy based on fundamental sound attributes of sound perception used for the PMS (pitch, loudness, tempo and "timbre"). Finally, it presents the results of the perceptual experiment guidance task with respect to these PMS strategies.

The obtained results would bring relevant information for the prediction of the user's behavior with a chosen sonification and constitute a first step toward general guidelines for mapping data onto display dimensions.

2 The concept of "target"

In several applications, the sound is used as a dynamic status and progress indicator in order to guide or inform the user on the evolution of a dynamic system. In such tasks, instead of directly monitoring the state of the system's parameter, it is possible to define a distance between the current and a desired state. The information that sonification needs to convey corresponds to one (or several) particular system's state(s) or to one point in an n-dimensional space in which the user (or any process) is evolving. These states, that may change over time, are defined as "targets".

In a driving aid application, for example, the target can represent the optimal speed for the road section on which the user is located. The system will then give information on the speed difference to be applied to the car to reach the optimal speed. The display information will be the distance between the current value and the target (e.g. the distance between the current speed and the optimal speed).

This concept of "target" allows to treat a number of specific application cases with a general sonification design. It is then possible to avoid the commonly encountered problem of the mapping of information to the display parameters (such as: "frequency is better to represent temperature than loudness") and to characterize different types of sonification strategies affected by the sound parameters (pitch, loudness, tempo, ...). This concept of "target" is, however, specific to inform or guide the user about a particular state and is difficult to apply to the simple monitoring of several state cases.

The aims of the sonification when guiding a user toward a target can be multiple:

- to guide as precisely as possible,
- to guide as quickly a possible,

– to guide without passing the target.

These guidance behaviors should be directly affected by the sound design choice as some acoustic properties may mainly affect the speed, and others the precision. To take into account these guidance behaviors, a taxonomy of sound strategies, based on sound morphology, is proposed in the next section.

3 Sound strategies

As mentioned in the introduction, several sound parameter may be used to convey data information. The use of the “target” notion allows the use of any of these parameters independently of a meaningful correlation of auditory dimensions and display dimensions. It is then possible to use any sound parameter to represent any display dimension but, in addition to the aesthetic problem, all parameters will not lead to the same level of information which may affect the guidance task in both precision and time. In order to characterize the ability of the sound parameters to convey information and guide the user, it is important to classify them.

This section aims at defining a taxonomy of sound strategies for sonification. It begins with the definition of this taxonomy then, after some considerations about the general approach, presents the different sound strategies designed in each category and tested in the experiment.

3.1 Towards a taxonomy of sound strategies

The proposed taxonomy is based on the definition of three morphological categories:

– Basic strategies: These strategies are based on the variation of basic perceptual sound attributes such as pitch, loudness, tempo or brightness. The sound attribute is directly a function of the distance to the target. Two polarities may be chosen (the attribute is maximum on the target or the attribute is minimum on the target). Furthermore, these strategies are constrained by human perceptual limits and the maximum reachable precision will be defined by the just noticeable difference of the attribute.
– Strategies with reference: The idea here is to include a sound reference corresponding to the target. Using this reference, the user should be able (by comparing the varying parameter to the reference) to evaluate the distance to the target without exploring the strategy. In the pitch case, adding a reference tone will produce modulation strength (if the frequencies are close) near the target and a continuous sound on the target. The same concept can be applied to the loudness using the concept of emergency of a masked sound or to the tempo with desynchronized rhythms. It is also possible to use an implicit perceptual reference such as the inharmonicity (the sound is harmonic only on the target) or the roughness (there is no roughness on the target).

- Strategies with reference and "zoom effect": In order to increase the precision around the target and to reduce the target's identification time, it is possible to augment the "strategies with reference" concept by adding a "zoom effect". This zoom consists in duplicating the strategy in different frequency bands. For example, in the case of the pitch with reference, rather than constructing the strategy with a pure tone, the use of a harmonic sound with several frequencies will create different modulations within different frequencies.

Giving these different sound strategy categories, it is possible to create a number of sonification strategies based on sound variations to evaluate the effect of the sound morphology on the user's response. Figure 1 shows the spectrograms of three sound strategies. For each strategy, the figure highlights the spectral evolution of the sound as function of the normalized distance (a distance equal to one corresponds to the maximum distance and a distance equal to zero corresponds to the target position).

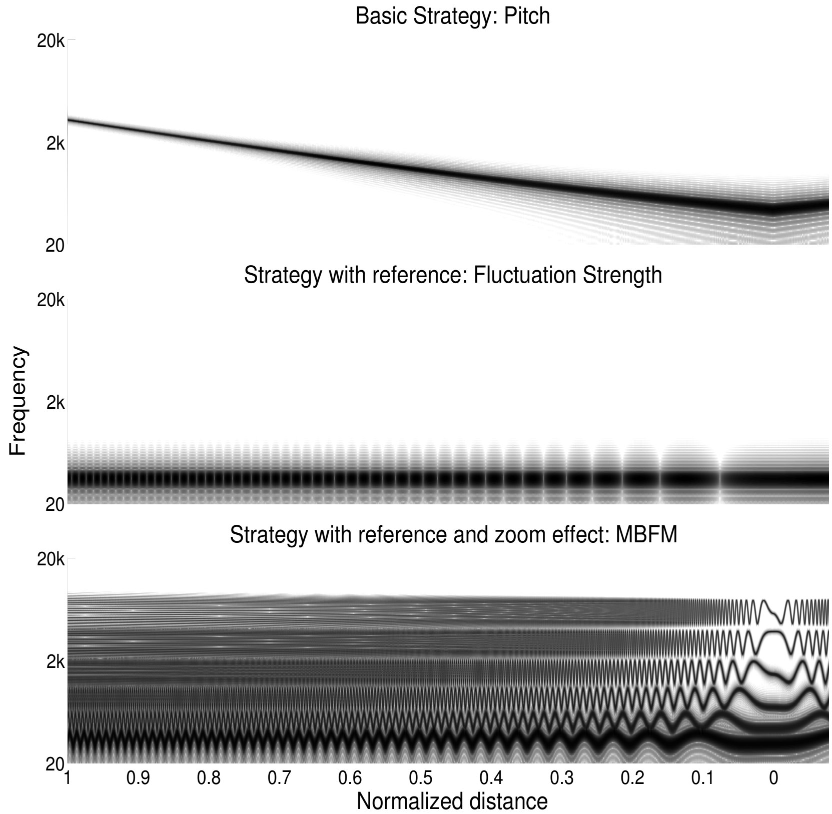

Fig. 1. Spectrograms of the sounds of three sonification strategies controlled by a varying normalized distance from 1 to 0 and back to 0.1. Pitch strategy on top, Fluctuation Strength in the middle and Multi-Band Frequency Modulation at the bottom.

We assume that basic strategies will be effective to guide the user quickly toward the target, but the user will need to pass the target in order to find the inflection point of the sound attribute (as the user is not familiar with the sound corresponding to the target beforehand). Strategies with reference should give higher precision than basic strategies but might take more time (the closer the target, the longer the time to perceive sound differences). Strategies with reference and zoom effect should give faster results and better precision than strategies with reference since high frequency components induce more rapid beats.

3.2 Designing of the sound strategies

In [18], Walker pointed out at least three design choices for the mapping sonification issue.

First, he highlighted the importance of the sound parameter choice to represent a given data dimension (e.g. pitch, loudness, tempo or timbre). Using the "target" concept, the mapping parameter choice is no longer important as the sound dimension is related to a normalized distance and not to a specific data dimension.

Second, according to Walker, designers must choose an appropriate polarity for the data-to-display mappings. Previous studies highlighted preferences for one or the other polarities for a particular application depending of the users' mental model. With the taxonomy introduced above this issue seems to be important for basic parameter strategies (such as pitch, loudness, tempo or brightness) but out of interest for strategies with references and strategies with reference and zoom effect. Furthermore, with the concept of target (which is not related to a specific data type), the notion of positive or negative polarity is difficult to establish and both polarities can be chosen. Considering that this problem is principally due to the user's perceptual association between data and sound dimension, we will not consider the potential influence of the mapping polarity of basic parameter strategies in this article.

The third design choice corresponds to the effect of the data dimension on the display dimension (how much change in sound parameter is needed to convey a given change in the data dimension?). For this scaling, the designer must choose a mapping function (linear, exponential, etc.) and a range of values. The mapping functions for this article were chosen to have the same precision of the sound parameter everywhere (perceptually linear). Defining a range of values for each sound parameter is still problematic as there is no definition of the Just Noticeable Difference for each sound parameter used in this study (apart from basic strategies). Aware that this choice may affect the results, we tried for each sound parameter to define the range of values that we believed was the most efficient.

On the basis of the proposed taxonomy, nine sound strategies were created while taking into account these design choices. These strategies are described below and sound examples resulting from these strategies controlled by a varying normalized distance from 1 to 0 and back to 0.5 are available online [1].

[1] `http://www.lma.cnrs-mrs.fr/~kronland/SonificationStrategies/`

Basic strategies

Pitch This strategy is based on frequency perception of sounds. It is constructed with a sinusoid with a varying frequency depending on the normalized distance between the user and the target:

$$s(t) = \cos(2\pi f(x)t)$$

with $x \in [0, 1]$, the normalized distance between the user and the target. Several scaling functions $f(x)$ can be considered (linear, logarithm, exponential, etc.). As human perception of frequency is logarithmic, we used the following function (weighted by the isophonic curve from the ISO 226 norm [1] to vary independently from the loudness) :

$$f(x) = f_{min}.2^{x.n_{oct}}$$

with $n_{oct} = \ln \frac{f_{max}}{f_{min}} \times \frac{1}{\ln 2}$ the number of octaves covered by the strategy and f_{min} and f_{max} the extreme values of the frequency.

The pitch polarity refers to the direction of pitch change as the user approaches the target:

- the frequency is minimal on the target ;
- the frequency is maximal on the target.

For this experiment, the polarity was chosen so that the frequency was minimal on the target. While the frequency perception limit is often quoted at 20 – 20000 Hz for young healthy listeners, the range of the scaling function was set to the frequencies corresponding to the traditional phone bandwidth (300 – 3400 Hz): $f_{min} = 300Hz$ and $f_{max} = 3394Hz$, hence spanning over 3.5 octaves.

Loudness This strategy is based on loudness perception of sounds. It is constructed with a sinusoid with a varying amplitude A depending on the normalized distance x between the user and the target:

$$s(t) = A(x).\cos(2\pi f_0 t)$$

As human perception of loudness is exponential, $A(x)$ is:

$$A(x) = 10^{[(\log A_{max} - \log A_{min}).x + \log A_{min}]}$$

For consumer applications (on mobile phone, for example), the maximal available level dynamic is around 40 dB. For the experiment, the polarity was chosen so that the loudness was minimal on the target: $a_{min} = -40dB$ and $a_{max} = 0dB$ so that $A_{min} = 10^{-\frac{a_{min}}{20}} = 0.01$ and $A_{max} = 1$.

Tempo This strategy is based on temporal perception of sounds. It is similar to the famous Geiger counter, often used as a sonification metaphor. This metaphor consists in repeating a stimulus and varying repetition rate. Thus the closer the target, the faster the sound repetition.

The sound stimulus used is a pulse tone of $f_0 = 1000$ Hz and $T = 0.1$ *sec*. The repetition rate is 20 Hz (1200 *bpm*) on the target and 2 Hz (120 *bpm*) for the maximum normalized distance $x = 1$.

Brightness This strategy is based on brightness perception of sounds. The brightness is considered to be one of the strongest perceptual attribute of the timbre of a sound. It corresponds to an indication of the amount of high-frequencies content in a sound, and is defined by the spectral centroid [3]. Brightness variations are obtained with a second order lowpass filtered white noise with a logarithmic distance dependent cutoff frequency F_c:

$$F_c(x) = f_{min}.2^{x.n_{oct}}$$

with $n_{oct} = \ln \frac{f_{max}}{f_{min}} \times \frac{1}{\ln 2}$ the number of octaves covered by the strategy and f_{min} and f_{max} the extreme frequency values.

As for the pitch strategy, the range of the scaling function was set to frequencies corresponding to the traditional phone bandwidth (300 – 3400 Hz): $f_{min} = 300Hz$ and $f_{max} = 3394Hz$ to cross 3.5 octave.

Strategies with reference

Fluctuation Strength This strategy is an extension of the pitch strategy. It uses a fixed tone as a reference for the target and a varying tone to inform about the normalized distance of the target. The first tone is a sinusoid with a frequency of $f_0 = X\ Hz$ (which is the reference), the second is a sinusoid with a frequency varying from $f_0 + 10\ Hz$ to f_0:

$$s(t) = 0.5 * cos(2\pi f_0 t) + 0.5 * cos(2\pi(f_0 + 10x)t)$$

The result is an amplitude modulation with a frequency equal to the difference between the two tones [8]. When the normalized distance x equal one, there are 10 modulations per second. When the target is reached, no more beats are heard.

Synchronicity This strategy is an extension of the tempo strategy. Two pulse tones are repeated. The first is the reference, the second is shifted with a varying time Δt depending on the distance between the user and the target. When the distance is maximum ($x = 1$), the second pulse is shifted by 1/4 of the pulsation frequency. When the target is reached, the two pulses are synchronized.

Inharmonicity This strategy is based on inharmonicity perception of sounds. It uses an implicit perceptual reference: the harmonic sound. It is constructed with a sum of sinusoids whose fundamental frequency is $f_0 = 200\ Hz$ and with higher frequencies computed with the piano' inharmonicity formula [22]:

$$s(t) = cos(2\pi f_0 t) + \sum_{k=2}^{N+1} \cos(2\pi f_k \sqrt{1 + b(x)\frac{f_k^2}{f_0^2}} t)$$

with $f_k = kf_0$, and $b(x)$ the inharmonicity factor: $b(x) = \frac{x}{100}$, that varies between 0 and 0.01.

Strategies with reference and zoom effect

Multi-Band Frequency Modulation (MBFM) This strategy is based on the frequency modulation of a harmonic sound (a sum of sinusoids with frequencies equal to integer multiples of a fundamental frequency). Instead of a conventional frequency modulation, each partial is here frequency modulated by a different modulating frequency : the higher the partial frequency, the higher the frequency of the modulating signal. When the user is getting closer to the target, the modulating frequencies decrease (there is no modulation when the target is reached). The farther the target, the higher the modulating frequency and the more complex the sound:

$$s(t) = \sum_{k=1}^{N} sin(2\pi f_k t + Ik.sin(2\pi f_m(x)t))$$

with f_k the frequency of the k^{th} harmonic, I the modulation index, and $f_m(x) = 10x$, the modulation frequency.

The use of a harmonic sound allows to construct an “auditory zoom”. The concept is simple: the frequency modulation affects all the harmonics with different temporalities. For a fixed distance, the higher the frequency, the faster the modulation. Near the target, the modulation frequency of the first harmonic is too small to rapidly grasp the target, but the modulation from the second harmonic which is twice as fast and then from the third harmonic (three times faster) allow to find the target faster and with more precision.

Multi Scale Beating (MSB) This strategy is based on the same concept as the MBFM strategy. It uses a sound of N harmonics M times duplicated and transposed by a few hertz with a factor that depends on the target distance. On the target the transposition factor is zero. The farther the target, the higher the transposition factor and the more complex the sound:

$$s(t) = \sum_{k=1}^{N} \sum_{m=0}^{M} A_k \cos(2\pi f_k(1 + m(\alpha(x) - 1))t)$$

with $\alpha(x) \in [0.94, 1.06]$ characterizing transposition factors. On the target ($\alpha(0) = 1$) $m(\alpha - 1)$ is equal to zero and different when moving away from the target. This creates a spectrum of M versions of N harmonics transposed by factors $m(\alpha - 1)$. It results in an “auditory zoom” due to the use of the harmonic sound and to the transposition factors that depend on the harmonic order.

4 Method

The first aim of the experiment was to evaluate the strategies’ capacity to dynamically guide a user toward a target in a one dimensional space with one polarity. Then the aim was to quantitatively assess the potential behavioural differences in the grasping induced by the different strategy' categories. The experiment was based on a within-subject design with a simple hand guidance task on a pen tablet.

4.1 Subjects

A total of 28 subjects participated in the experiment (8 women and 20 men). Mean age: 33 ± 12 years (min. 21; max. 58 years); 21 subjects were considered as experts (musicians or people working in the audio field), 7 were non-experts. All were naive regarding the purpose of the experiment. No audiogram was performed, but none of the subjects reported hearing losses.

4.2 Stimuli and Apparatus

The subjects were equipped with a stereo closed-ear headphone (model Sennheiser HD280). They were placed in a quiet room, in front of a pen tablet with a 24" screen display (model Wacom Cintiq 24HD). The experimental session was run using a Max/MSP interface running on an Apple computer. The stimuli were synthesized in real-time with the nine strategies defined in section 3.2 as function of the normalized distance between the pen and the target.

4.3 Procedure

The nine sonification strategies were evaluated with a guiding task on eight distances (10, 12.5, 15, 17.5, 20, 22.5, 25, and 27.5 cm with a maximum distance to the target of 30 cm, that gives normalized distances of: 0.33, 0.47, 0.5, 0.58, 0.67, 0.75, 0.83, and 0.92). The experiment was divided in nine blocs of nine trials, each bloc containing the nine strategies in a random order. Participants were told to find, for each trial, a hidden target randomly placed on a line. They were not given any explanation about each sound strategy, and were just instructed that the real-time sound transformations would inform them about the distance between the pen and the hidden target. The first bloc was considered as a training bloc and the subjects were free to explore the sound transformations without instruction. After this training, they were instructed to find the target as quickly, as accurately, and as directly as possible without removing the pen from the tablet. In each trial, subjects first were to place the pen on a starting position (on the left side of the screen). Then, they started the trial by pushing the space bar of a keyboard and moving on a visual line to find the position of the hidden target. Finally, when the target was identified, the subjects validated the estimated position of the target with the space bar of the keyboard to proceed to the next trial. The order of the trial presentation was calculated so that a strategy never appeared twice in a row and a distance never appeared three times in a row. The order and the starting position were randomized to avoid any potential learning effect and to exclude any possible visual aid on the screen. No accuracy feedback was provided after each trial.

5 Results

Figure 2 shows typical trajectories collected on several subjects during a trial. The analysis of these trajectories highlights different subject's behaviors to find the target. First, the total duration of the task performance was spread between a few seconds to 40 *sec*. The subjects were able to find the target directly in some cases (figure 2 at top left) or with a lot of direction changes in other cases (figure 2 at top right). Finally the error can vary from less than 1 cm to more than 15 cm (figure 2 at bottom right).

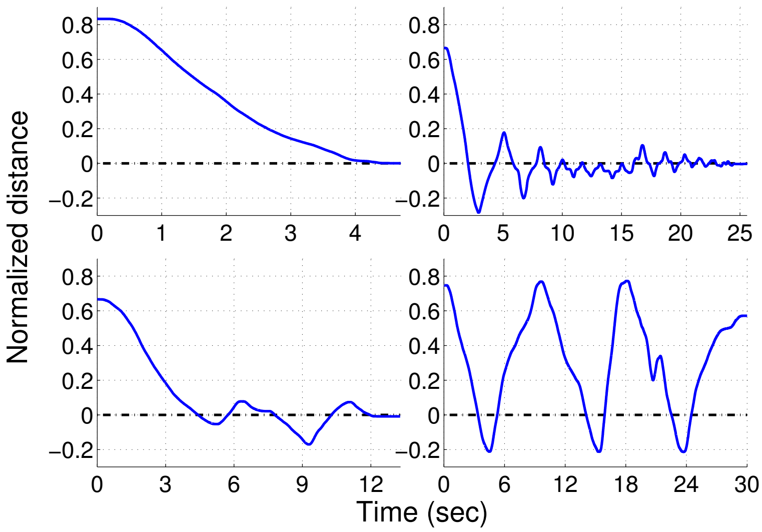

Fig. 2. Different examples of distance evolution to find the hidden target.

Although it would be interesting to analyse and classify these trajectory types regarding the subjects and the sound strategies, this article analysed the effect of each strategy on the guiding task only by comparing the final precision, the total time, and the number of oscillations needed to find the target. The analysis was performed by separating expert (musician or people working on audio processing) and non-expert subjects.

5.1 Precision

The final precision corresponds to the distance between the pen and the target when the subject validated the trial. The mean value of the final precision with the 95 % confident interval is represented in figure 3 for non expert (top) and expert subjects (bottom). There is a large difference between these two groups. The mean value of final precision for the non-expert group is 4.58 ± 2.88 % while it is 1.04 ± 0.75 % for the expert group. The final precision as function of the strategy highlights a strong effect of the strategy. The bests strategies for the non expert group are the *Tempo* and the *Inharmonicity* while the worst

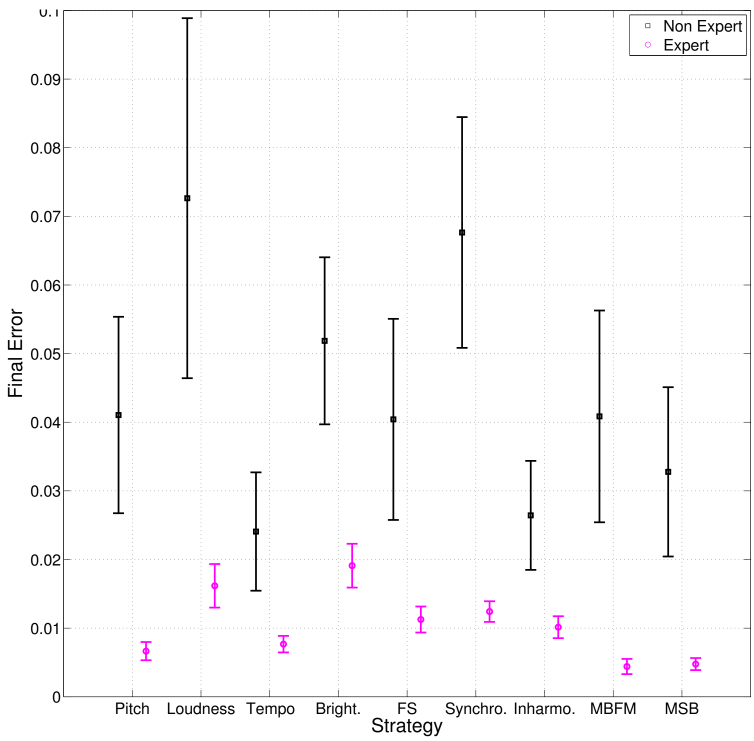

Fig. 3. Mean and 95 % confident interval of the final precision for non expert (top) and expert (bottom) subjects as function of the sound strategy.

are the *Loudness* and the *Synchronicity*. For the expert group, the best results are obtained with the reference and zoom effect strategies (the *MBFM* and the *MSB*) and the worst are obtained with the *Loudness* and the *Brightness* strategies. For the expert group, the results for the three strategies with reference are similar and the results for the two strategies with reference and loop are equal. The results for simple parameter strategies show large differences between *Pitch* and *Tempo* in one part and *Loudness* and *Brightness* in another part. This can be explained by the perceptual resolution of these parameters that is high for pitch and tempo perception and low for loudness and brightness perception.

5.2 Time

The mean of total time spent for all the experiment is 30 ± 10 *min* The mean of total time spent in each trial with the 95 % confident interval is represented in figure 4 for non expert (square) and expert subjects (circles). The mean response time for one trial is 11.2 ± 3.2 *sec* for the non expert group and 12.9 ± 2.6 *sec* for the expert group. If the expert subjects spend more time on each trial, the smaller standard deviation shows more consistency across all the trials.

If the results tend to show the same tendencies for both groups, the mean time spent on *Tempo* and *MBFM* strategies is significantly shorter for the non expert group. Despite a strong effect of the strategy on the mean response time, it is difficult to find an effect of the morphological category types. A shorter

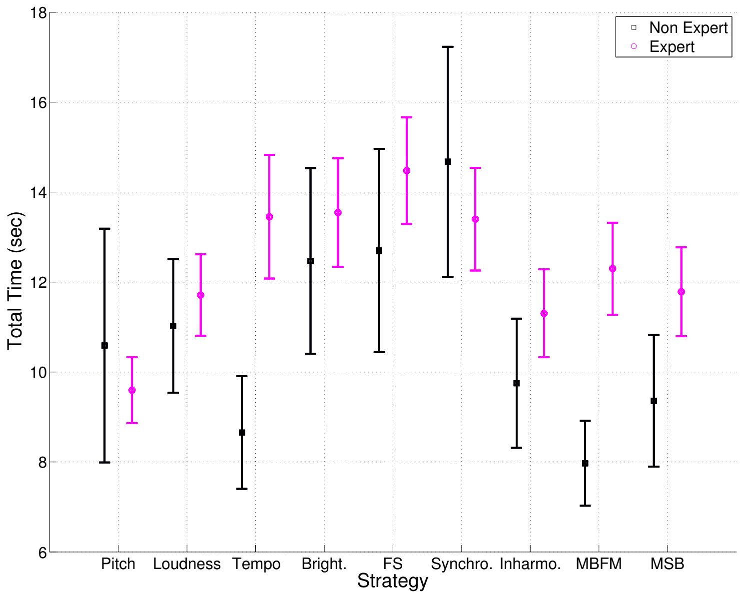

Fig. 4. Mean and 95 % confident interval of the total time for one trial for non expert (top) and expert (bottom) subjects as function of the sound strategy.

response time was expected for simple parameter strategies, while it turns out that this is only true for the *Pitch* strategy for the experts and for the *Pitch* and *Tempo* strategies for the non experts. As expected, the two groups spent most time on strategies with references (*Fluctuation Strength* and *Synchronicity*) because of the sound morphology of these strategies.

5.3 Number of oscillations

The number of oscillations around the target corresponds to the number of direction changes around the target. On the examples of figure 2, the number of oscillations is 0 for the first example (top-left), 24 for the second (top-right), 4 for the third (bottom-left) and 5 for the fourth (bottom-right). Regarding these examples, it seems clear that these oscillations don't represent the same pointing behavior for each trial. For the second trial, the oscillations are close to the target and allow the subject to refine the perceived position. For the fourth trial, the oscillations have large amplitudes and no convergence tendency, which highlights a lack of strategy understanding.

The mean number of oscillations for each trial with the 95 % confident interval is represented in figure 5 for non expert (square) and expert subjects (circles). Expert subjects made more oscillations around the target than non expert subjects (6.3 ± 2.0 vs. 3.6 ± 1.5). Except for the *Tempo*, the expert group made more oscillations (between six and seven) for the basic parameter strategies than for the other groups of strategies. The strategies with reference parameters led to the same number of oscillations (e.g. five) and the two strategies with reference and

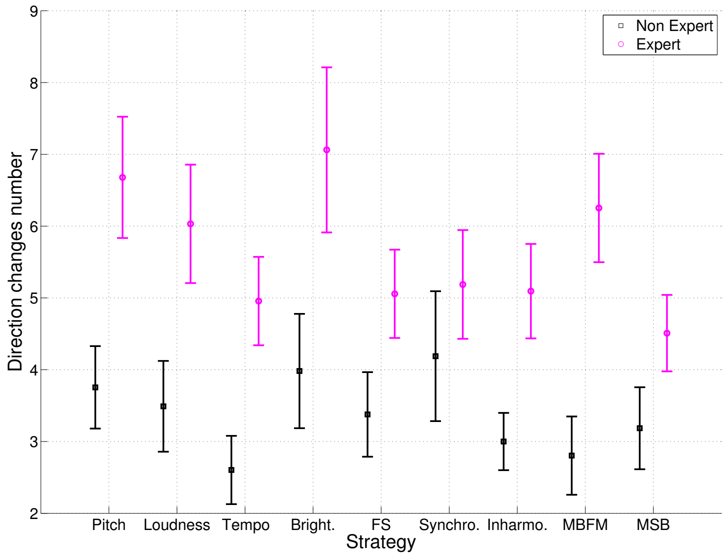

Fig. 5. Mean and 95 % confident interval of the number of oscillations around the target for one trial for non expert (top) and expert (bottom) subjects as function of the sound strategy.

zoom led to different results (a mean of six oscillations for the *MBFM* strategy and of 4.5 oscillations for the *MSB* strategy).

5.4 Discussion

The purpose of this study was to evaluate the capacity of various sound strategies to guide a user toward a target and to explore the potential behavior differences induced by the different morphological categories.

The results first show a great difference between expert and non-expert subjects with better results for expert. This difference especially appears on the final precision and seems to be due to the listening expertise of the subjects. On the other side expert subjects spend more time on each trial and make more oscillations around the target than non-expert subjects.

Both subject groups succeeded in finding the target with less than 5 % of error within less than 15 seconds confirming the capacity of the proposed strategies to properly guide the user toward the target. Result comparisons as function of the sound strategies highlight an influence of the sound parameters on the user behavior in the grasping movement. For example, expert subjects were more precise using MBFM and MSB strategies than with Pitch strategy, but they were also slower. In general, both groups showed difficulties with Loudness and Brightness strategies, that seem to be the worst strategies in terms of precision. Surprisingly, non-expert subjects have better performances with the inharmonicity strategy than with the pitch strategy, which validates the implicit perceptual reference hypothesis on this factor.

The search for similarities within each strategy category (*basic*, *with reference* and *with reference and zoom effect*) has so far been unsuccessful with respect to the final precision, the total time and the number of oscillations around the target. For example, in basic strategies, pitch and tempo led to significantly better results than loudness and brightness strategies. It therefore seems necessary to apply a more complex analysis than in the present paper to evaluate each sound strategy.

6 Conclusions

In the current study, a new parameter mapping sonification design approach was introduced. This approach is first based on the definition of a "target" concept which aim is to facilitate the separation between the informative sound strategy and the information to display. This separation allows to evaluate given sound strategies independently from the application and to predict the sonification result for any type of guiding task. We expect that the use of a normalized distance to the target allows to use different strategy types for the same data type. In this article, a taxonomy of sound strategies based on sound morphology was proposed. The aim of this taxonomy was to predict the user's guidance behavior with the use of several sound strategies. An experiment based on a guidance task was realized in order to evaluate the user's guidance behavior. The results highlighted great differences between expert and non expert users and showed the influence of the sound strategies on the guidance behaviors. While some sound strategies allow to quickly guide the user towards the targets other strategies may allow a better precision or guide more directly to the target. Nevertheless, a simple result analysis does not allow to link the guidance behaviors to the three defined morphological categories. Therefore, it would be interesting to analyze the guidance behaviors with a more complex analysis of the grasping movement toward the targets.

This study only focused on the user evaluation of morphological categories. Therefore it is not possible to conclude on the efficiency of the "target" concept in parameter mapping sonification. It would be interesting to test this concept with an experiment based on the sonification of different applications so as to verify the stability of the guidance behavior toward different types of normalized distances.

References

1. Normal equal-loudness level contours - ISO 226. Acoustics International Organization for Standardization, 2003.
2. S. Barrass. Personify: A toolkit for perceptually meaningful sonification. In *ACMA*, 1995.
3. J. W. Beauchamp. Synthesis by spectral amplitude and" brightness" matching of analyzed musical instrument tones. *Journal of the Audio Engineering Society*, 30(6):396–406, 1982.

4. O. Ben-Tal, J. Berger, B. Cook, M. Daniels, G. Scavone, and P. Cook. Sonart: The sonification application research toolbox. In *Proceedings of the 8th International Conference on Auditory Display, Kyoto, Japan*, 2002.
5. J. W. Bruce and N. T. Palmer. Sift: Sonification integrable flexible toolkit. In *Proceedings of the 11th International Conference on Auditory Display, Limerick, Ireland*, 2005.
6. R. M. Candey, A. M. Schertenleib, and W. L. Diaz Merced. xsonify: Sonification tool for space physics. In *Proceedings of the 12th International Conference on Auditory Display, London, UK*, 2006.
7. B. K. Davison and B. N. Walker. Sonification sandbox reconstruction: Software standard for auditory graphs. In *Proceedings of the 13th International Conference on Auditory Display, Montréal, Canada*, 2007.
8. H. Fastl and E. Zwicker. *Psychoacoustics: Facts and Models*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006.
9. S. Ferguson, D. Cabrera, K. Beilharz, and H.-J. Song. Using psychoacoustical models for information sonification. In *Proceedings of the 12th International Conference on Auditory Display, London, UK*, 2006.
10. F. Grond and J. Berger. Parameter mapping sonification. In T. Hermann, A. Hunt, and J. G. Neuhoff, editors, *The Sonification Handbook*. Logos Publishing House, 2011.
11. T. Hermann, A. Hunt, and J. Neuhoff, editors. *The Sonification Handbook*. Logos Publishing House, Berlin, Germany, 2011.
12. A. J. Joseph and S. K. Lodha. Musart: Musical audio transfer function real-time toolkit. In *Proceedings of the 8th International Conference on Auditory Display, Kyoto, Japan*, 2002.
13. G. Kramer. *Some organizing principles for representing data with sound*. Addison Wesley Longman, 1992.
14. G. Kramer. *Auditory Display: Sonification, Audification and Auditory Interfaces*. Perseus Publishing, 1993.
15. S. K. Lodha, J. Beahan, T. Heppe, A. Joseph, and B. Zane-Ulman. Muse: A musical data sonification toolkit. In *Proceedings of the 4th International Conference on Auditory Display, Palo Alto, California*, 1997.
16. S. Pauletto and A. Hunt. A toolkit for interactive sonification. In *Proceedings of the 10th International Conference on Auditory Display, Sidney, Australia*, 2004.
17. A. M. Schertenleib. xsonify, 2005.
18. B. N. Walker. Magnitude estimation of conceptual data dimensions for use in sonification. *J. of Experimental Psychology Applied*, 8(4):211–221, 2002.
19. B. N. Walker. Consistency of magnitude estimations with conceptual data dimensions used for sonification. *Applied Cognitive Psychology*, 21(5):579–599, 2007.
20. B. N. Walker and G. Kramer. Mappings and metaphors in auditory displays: An experimental assessment. In *Proceedings of the 3rd International Conference on Auditory Display, Palo Alto, California*, 1996.
21. B. N. Walker and G. Kramer. Mappings and metaphors in auditory displays: An experimental assessment. *ACM Transactions on Applied Perception*, 2(4):407–412, October 2005.
22. R. W. Young. Inharmonicity of plain wire piano strings. *Journal of the Acoustical Society of America*, 24(3):267–273, May 1952.